Journal of Computer Science and Frontier Technologies
ISSN Print: 3104-4204
ISSN Online: 3104-4212
CD ISSN: 3104-4220
https://doi.org/10.63313/JCSFT.9014

A High-Precision Object Detection Algorithm for UAV Aerial Images: WI-YOLO

Liu Yang
School of Electronic Engineering, Tianjin University of Technology and Education, Tianjin, 300222, China
Email: 2894076249@qq.com

How to cite this paper: Liu, Y. (2025). A High-Precision Object Detection Algorithm for UAV Aerial Images: WI-YOLO. Journal of Computer Science and Frontier Technologies, 1(2), 39-51. ISSN Print: 3104-4204; ISSN Online: 3104-4212.
https://doi.org/10.63313/JCSFT.9014
Published: 2025-10-16

Abstract

To tackle the challenges of low small-target detection accuracy, intense background interference, and high model computational complexity in unmanned aerial vehicle (UAV) aerial images, this study presents an enhanced lightweight object detection algorithm, namely WI-YOLO. Built on the YOLOv11n architecture, the algorithm integrates two key improvements: first, an improved efficient multi-scale attention mechanism (iEMA) is embedded into the backbone network, which strengthens the discriminability of low-contrast targets by fusing standard deviation and local contrast features; second, a wavelet transform convolution module (WTConv) is designed to expand the receptive field and preserve multi-scale structural information via a multi-level frequency-domain decomposition and reconstruction mechanism.

Experimental evaluations on the DOTA dataset show that WI-YOLO achieves 61.5% in mAP@50 and 35.1% in mAP@50:95—marking significant gains over the baseline model. Concurrently, the model’s parameter count is reduced to 5.9 MB, and its inference speed is increased to 30.5 FPS. These results demonstrate the coordinated optimization of detection accuracy and computational efficiency, making WI-YOLO well-suited for real-time detection tasks on UAV embedded platforms.

Keywords

UAV Aerial Images; Object Detection; YOLOv11; Attention Mechanism

1. Introduction

With the extensive integration of computer vision and unmanned aerial vehicle (UAV) platforms, object detection technology based on aerial images has played a crucial role in numerous fields, such as smart city management, agriculture and forestry condition monitoring, emergency disaster response, and national defense security patrols. However, aerial images are affected by multiple factors, including UAV flight altitude, sensor perspective, atmospheric light variations, and environmental disturbances. As a result, these images often exhibit inherent characteristics:

significant background interference, wide distribution of target scales, dense small target instances, widespread local occlusion, and easy loss of detailed information. Collectively, these factors lead to problems of insufficient recognition accuracy and high missed detection rates for general detection models in practical aerial scenarios. Traditional detection methods mostly rely on image stitching strategies and manually designed feature extractors, which have limited representational capabilities in complex backgrounds and struggle to achieve an effective balance between detection accuracy and system real-time performance. Therefore, in response to the urgent demand for efficient perception technology in practical UAV applications, there is an urgent need to develop detection algorithms that can simultaneously handle multi-scale targets while balancing high accuracy and lightweight design.

Currently, object detection algorithms are mainly categorized into two types based on their implementation mechanisms: two-stage methods based on region proposals and one-stage methods based on end-to-end regression. Two-stage detection algorithms (e.g., Faster R-CNN and Mask R-CNN) first generate region proposals, followed by performing classification and bounding box regression on these regions. Although such methods have advantages in detection accuracy, their inference speed is typically limited due to the step-by-step execution of the process. In contrast, one-stage algorithms represented by the SSD (Single Shot Multibox Detector) and YOLO (You Only Look Once) series adopt a unified convolutional network to directly predict target categories and positions in images, eliminating the region proposal generation step. They can output detection results with a single forward propagation, thus demonstrating more excellent performance in detection efficiency. Yi et al. proposed the LAR-YOLOv8 algorithm based on YOLOv8: they enhanced local modules via a dual-branch attention mechanism and introduced the RIOU loss function to improve the shape consistency between predicted boxes and ground truth boxes, effectively addressing the limitations of small target detection in remote sensing images. Zhu et al. proposed the generalized oversampling Prouhet-Thue-Morse (OGPTM) method, which integrates a pixel-wise multiplication processor (PmuP) to achieve better sidelobe suppression performance in high-speed object detection. Xie et al. proposed FARP-Net for 3D point cloud object detection tasks, innovatively introducing a weighted relation-aware proposal module (WRPM) to overcome the limitations of traditional context information extraction methods. Zhang Huawei et al. improved the activation function of the attention gate and proposed the GUS-YOLO algorithm based on the YOLO series, which enhances object detection performance on remote sensing datasets.

Despite significant progress made by the academic community in improving the performance of object detection for UAV aerial images, existing methods still face three key challenges: high missed detection rates of small targets in complex backgrounds, unbalanced detection accuracy for multi-scale targets, and difficulty in balancing model computational efficiency and detection accuracy. To address these

issues, this paper proposes an improved algorithm, WI-YOLO, based on the YOLOv11n architecture. Through a modular design, this algorithm not only improves the detection accuracy of small targets but also significantly reduces model complexity and optimizes the utilization of computing resources, providing an efficient and reliable detection solution for UAV aerial scenarios. The main contributions of this paper are as follows:

(1) An improved efficient multi-scale attention mechanism (iEMA) is designed. By introducing standard deviation and local contrast metrics, it enhances the model's feature discrimination ability in complex backgrounds and effectively improves the detection accuracy of low-contrast small targets.

(2) A wavelet transform-based convolutional module (WTConv) is proposed. Through a multi-level frequency domain decomposition and reconstruction mechanism, it expands the receptive field while retaining key frequency features, significantly improving the model's adaptability to multi-scale targets and realizing the lightweight design of the model.

2. Design of the WI-YOLO Algorithm

As the latest iteration of the YOLO series, YOLOv11 was released by the Ultralytics team in 2024 and quickly established a new state-of-the-art (SOTA) performance benchmark in the field of single-stage object detection. This architecture upholds the core idea of the YOLO series, which abstracts object detection as a single regression problem. It enables one-time inference on input images and simultaneously outputs bounding box coordinates and category probabilities, thereby maintaining high frame rates for real-time detection even in complex scenarios. To adapt to diverse deployment environments ranging from embedded devices to server clusters, YOLOv11 is categorized into five variants (n, s, m, l, x) based on model capacity and computational complexity. Its overall network inherits the classic modular composition: the Backbone (responsible for basic feature extraction), the Neck (for multi-scale feature fusion), and the Head (for final prediction) work synergistically. Together, they form the structural foundation for achieving an excellent balance between speed and accuracy.

Nevertheless, when YOLOv11 is directly applied to the specific field of UAV aerial image detection, its inherent design still shows shortcomings in addressing challenges such as small targets, complex backgrounds, and multi-scale variations. First, due to the generally low pixel coverage of targets from the aerial perspective, their subtle appearance features are easily overwhelmed during the model's forward propagation, leading to inaccurate localization and missed detections. Second, complex background interferences (e.g., ground scenes and vegetation shadows) make it difficult for the model to focus on effective target features, increasing the risk of false detections. Additionally, while the continuous downsampling operations in the deep network pursue high-level semantic information, they inevitably cause severe

loss of key details of small targets—this directly restricts the model's overall detection performance for targets with significant scale differences.

To address the aforementioned issues, this study proposes a small-target detection algorithm (WI-YOLO) for UAV aerial images, based on the lightweight YOLOv11n variant; its structure is illustrated in **Figure 1**. The algorithm primarily introduces two improvements: (1) integrating the iEMA attention mechanism into the Backbone to enhance the model's ability to extract features of small targets in complex backgrounds; (2) replacing some standard convolutions with the wavelet transform convolution module (WTConv) to improve the richness of feature expression and nonlinear fitting capability, while reducing the model's parameter count. Specifically, in the improved model, the residual blocks in the C3k2 module of the Backbone are replaced with iEMA modules, and some traditional convolutions are replaced with WTConv modules. Through these structural optimizations, the algorithm maintains high inference speed while significantly enhancing the detection performance for small targets in UAV images.

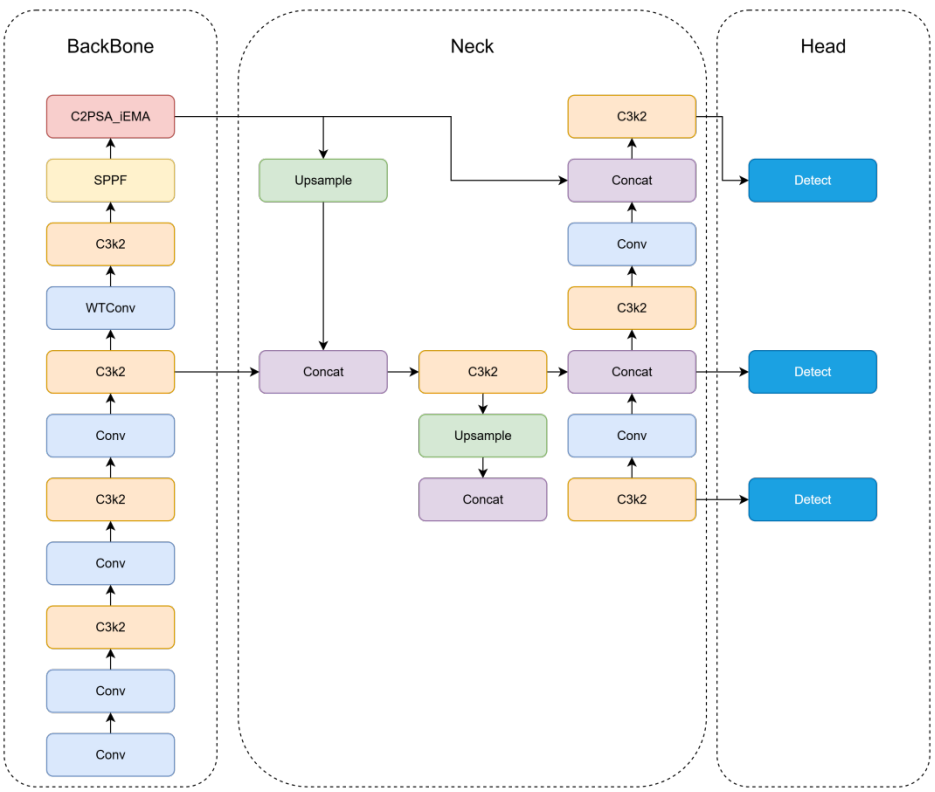

Figure 1. Structure of the WI-YOLO Model.

2.1. The iEMA Attention Mechanism

In the field of computer vision, the attention mechanism has emerged as one of the key technologies for improving model performance by simulating human cognitive processes and guiding computing resources to focus on critical regions in images.

Particularly in UAV-based object detection tasks, when convolution kernels slide over images to extract features, ubiquitous background information often interferes with the representational learning of real targets. The introduction of the attention mechanism can dynamically suppress these redundant background responses and enhance the activation of features related to target regions, thereby improving the quality of feature maps.

The Efficient Multi-Scale Attention (EMA) module adopts strategies such as feature grouping, parallel subnetwork construction, and cross-spatial dimension learning. While maintaining the integrity of information across channels, it significantly reduces the overall computational burden of the model. To further improve the numerical stability of this module and prevent numerical overflow caused by excessively large input values, this study replaces the Softmax activation function in the original EMA structure with the Log-Softmax function. This improvement ensures robust model performance while effectively accelerating the forward inference process, and the resulting module is defined as the improved Efficient Multi-Scale Attention (iEMA).

In the YOLOv11n architecture, the core responsibility of the Backbone is to perform hierarchical feature extraction. Among its components, the SPPF (Spatial Pyramid Pooling - Fast) module fuses contextual information from different receptive fields through multi-branch pooling operations, which is crucial for enhancing the model's multi-scale understanding ability. For this reason, this study embeds the iEMA module into the Backbone, placing it before the SPPF module. This design aims to use the cross-spatial attention of iEMA to pre-screen and enhance features, enabling it to form functional complementarity with the subsequent multi-scale fusion strategy of SPPF. Together, they enhance the model's detection capability, especially for small targets in UAV aerial images.

The traditional EMA mechanism mainly relies on mean statistics to aggregate features, which has limitations in processing low-contrast aerial images and cannot fully characterize the subtle differences between targets and backgrounds. To address this issue, this study introduces key enhancements to iEMA: on the basis of the original mean calculation, it simultaneously incorporates the metrics of standard deviation and local contrast. The standard deviation can effectively reflect the fluctuation of pixels in a region, helping to capture texture details; while local contrast strengthens the model's perception of edges and structural mutations. These two supplements enable the model to more accurately locate information-rich regions, significantly improving the feature discrimination ability for low-visibility targets and the overall robustness of the model.

2.2. The WTConv Wavelet Transform Convolution

In object detection models based on convolutional neural networks (CNNs), the input region range corresponding to each neuron on the feature map (i.e., the receptive field) is a key factor determining the model's ability to understand context. A

larger receptive field helps the model capture more extensive global information, thereby improving its ability to infer relationships between targets in complex scenarios. However, the standard convolution operation used in YOLO series models is inherently a local operation. Structurally, it is difficult to break through the limitations of local neighborhoods and achieve true large-scale information interaction. Existing studies have attempted to simulate the global attention mechanism in vision transformers by directly increasing the convolution kernel size, but this method often leads to a quadratic increase in computational complexity and only achieves limited performance improvement.

Traditional CNNs usually face dual challenges of a sharp increase in parameter count and a bottleneck in performance improvement when expanding the receptive field. More importantly, standard convolution operations inherently tend to extract high-frequency texture features from images, while they have obvious deficiencies in capturing low-frequency information critical for target recognition, such as overall structure, contour shape, and deep contextual semantics.

To solve the above problems, this study introduces a convolution module based on wavelet transform, namely WTConv. The core idea of this module is to use multi-level wavelet transform to decompose the input feature map into subband components of different frequencies. Subsequently, lightweight grouped depthwise convolution is implemented in the decomposed multi-scale frequency domain space. Finally, the processed multi-frequency band information is efficiently fused and reconstructed through inverse wavelet transform. While retaining the advantage of local modeling of traditional convolution, WTConv significantly enhances the model's ability to perceive multi-frequency structural information of images through frequency domain analysis and fusion. This enables it to synergistically utilize the target shape information contained in the low-frequency part and the detailed edge features retained in the high-frequency part in UAV aerial image detection.

To integrate the advantages of WTConv into the model, this study embeds it into the C3k2 module of the YOLOv11n Backbone, constructing an improved C3k2_WTConv module (its structure is shown in **Figure 2**). On the basis of retaining the efficient stacked design of the original C3k2, the improved module effectively expands the effective receptive field of the convolution layer through WTConv without significantly increasing the number of parameters. This enables the model to exhibit stronger feature capture capability, higher detection robustness, and higher recall rate in UAV aerial scenarios with variable scales and complex backgrounds, especially for small-sized targets.

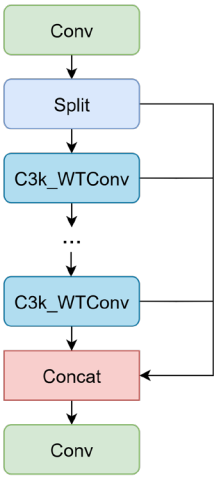

Figure 2. Structure of the iEMA Module.

3. Experiments and Results Analysis

3.1 Dataset

In this study, the DOTA dataset is adopted for the training and performance validation of the proposed algorithm. As a large-scale benchmark dataset specifically designed for small-scale object detection in aerial images, DOTA collects images from various sensors and aerial platforms, covering diverse terrain scenes and imaging conditions. In small object detection tasks, this dataset leads among similar aerial datasets in terms of both sample size and annotation quality, thus serving as the foundation of experimental data in this study.

The dataset contains a total of 3065 high-resolution aerial images, which are divided into the training set, validation set, and test set with the number of images being 1903, 629, and 533 respectively. All images fully retain the original 15 categories of target annotations from DOTA, including typical categories such as aircraft, vehicles, and ships, with the total number of annotated instances exceeding 180,000. The dataset covers various typical environments including cities, rural areas, and coasts, and the targets exhibit complex characteristics such as significant scale differences, dense spatial distribution, and variable orientations. These features enable sufficient evaluation of the algorithm’s generalization performance in real aerial environments.

Given that small target instances account for more than 60% of the dataset, this study introduces the Mosaic data augmentation method. By randomly concatenating four training images, this method effectively increases the occurrence frequency of small targets during the training process. Meanwhile, strategies such as random rotation, multi-scale transformation, and color space perturbation are employed to enhance the model’s adaptability to scale changes and light fluctuations. To adapt to the training process of YOLO series algorithms, the original oriented bounding

box (OBB) annotations are converted into the horizontal detection box format, and the prior anchor box sizes are regenerated based on the K-means clustering method. All input images are standardized to ensure numerical stability during the training process. The details of the dataset are illustrated in **Figure 3.**

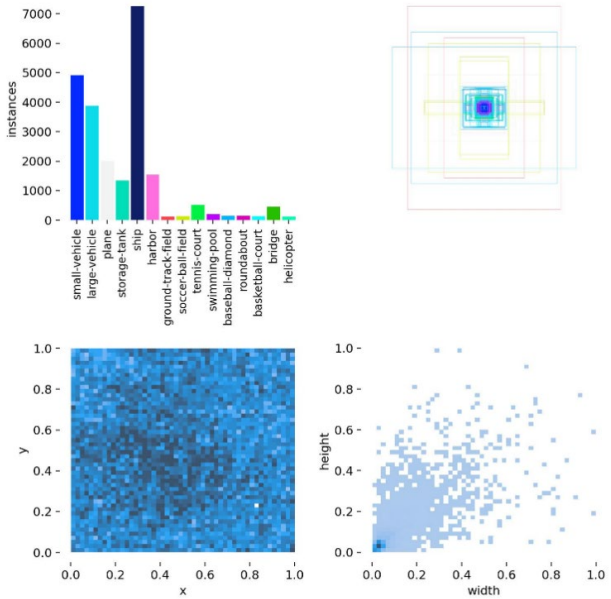

Figure 3. Dataset Visualization Distribution Map.

3.2. Experimental Environment and Evaluation Metrics

All experiments were conducted in the same hardware environment, with the relevant configuration details presented in **Table 1**.

Table 1. Table of Relevant Environment Configuration.

software and hardware configuration	version or model
CPU	Intel Xeon Gold 6226R
GPU	NVIDIA GeForce RTX 3080
Python	3.8.19
Pytorch	1.13.1
YOLOv11 version	8.3.24

In this study, common evaluation metrics for object detection tasks were employed to assess the algorithm’s performance, including mean average precision (mAP), recall (R), precision (P), model size, and floating-point operations per second (FLOPS).

Specifically, precision (P) reflects the proportion of actually positive samples among the samples predicted as positive by the model, while recall (R) represents the proportion of actually positive samples that are correctly predicted as positive by the model. These two metrics are defined based on three core statistical indicators: True Positives (TP), False Positives (FP), and False Negatives (FN). Among them, TP

denotes the number of samples predicted as positive and actually positive; FP refers to the number of samples predicted as positive but actually negative; FN represents the number of samples actually positive but predicted as negative.

Average Precision (AP) is calculated as the mean of precision values across different categories, which is used to evaluate the model's performance on a single category. Mean Average Precision (mAP), as the average of AP values across all categories, is employed to measure the model's comprehensive performance across all target categories.

The corresponding calculation formulas are shown in Equations (1), (2), (3), and (4):

$$R = \frac{TP}{TP+FN}. \tag{1}$$

$$P = \frac{TP}{TP+FP}. \tag{2}$$

$$AP = \int_0^1 P(R)dR. \tag{3}$$

$$mAP = \frac{1}{n}\sum_1^n AP_i. \tag{4}$$

3.3. Ablation Experiment

To systematically evaluate the contribution of each improvement strategy proposed in this study to model performance, ablation experiments were conducted using YOLOv11n as the baseline model. All experiments were performed under unified training settings, with different improvement modules introduced incrementally. The core evaluation metrics included precision (P), recall (R), mAP@0.5, mAP@50:95, FPS, and model size. Detailed comparison results are summarized in **Table 2**.

Table 2. Ablation Experiment Results.

Baseline	iEMA	EUCB	P/%	R/%	mAP @50/%	mAP @50:95/%	Model Size/MB	FPS
Yolov11n			65.6	56.6	56.7	32.6	9.3	26.6
	✅		66.0	57.1	56.9	33.6	8.4	28.2
		✅	65.9	58.8	59.0	32.9	6.3	27.7
	✅	✅	67.6	59.0	61.5	35.1	5.9	30.5

The ablation experiment results in Table 2 clearly demonstrate the contribution of each improvement module to the performance of the WI-YOLO model. In terms of detection accuracy, the complete WI-YOLO model (incorporating both iEMA and WTConv) achieved 61.5% in mAP@50 and 35.1% in mAP@50:95, which were 4.8 and 2.5 percentage points higher than those of the baseline model (YOLOv11n), respectively. Meanwhile, the precision (P) and recall (R) increased to 67.6% and 59.0%, respectively—indicating that the model not only maintains high classification accuracy but also significantly enhances its ability to identify positive samples.

In terms of model efficiency, WI-YOLO exhibits excellent lightweight characteristics. The parameter count of the complete model was reduced to 5.9 MB, which is a 36.6% decrease compared to the baseline model's 9.3 MB. Notably, the inference speed of

the model increased instead of decreasing, with FPS rising from 26.6 (baseline) to 30.5. This achieved the dual improvement of accuracy and speed.

An analysis of the independent contributions of each module reveals that the iEMA module mainly improved the model's precision and mAP@50:95, while compressing the model size to 8.4 MB. In contrast, the WTConv module significantly enhanced the recall and further reduced the model size to 6.3 MB. When the two modules work synergistically, they not only inherit their respective advantages but also produce a "1+1>2" effect, achieving optimal performance across all metrics. This demonstrates that iEMA and WTConv have good functional complementarity: the former enhances feature discrimination ability through the attention mechanism, while the latter optimizes feature extraction efficiency via wavelet transform.

In conclusion, WI-YOLO significantly reduces model complexity while comprehensively improving detection performance. Particularly, it achieves a significant accuracy improvement while maintaining high inference speed, providing an efficient solution for real-time object detection on UAV platforms.

3.4. Comparative Experiments

To verify the effectiveness of the proposed WI-YOLO model, comparative experiments were conducted: first, WI-YOLO was trained and validated on the dataset alongside mainstream object detection models (including YOLOv5n, YOLOv8n, YOLOv10n, YOLOv11n, etc.). Key metrics were compared across all models, including mAP@50, mAP@50:95, precision (P), recall (R), FPS, and model size. The results of these comparative experiments are presented in **Table 3**.

Table 3. Comparative Experiment Results.

Model	P/%	R/%	mAP @50/%	mAP @50:95/%	Model Size/MB	FPS
YOLOv5n	62.0	46.8	39.6	24.3	3.6	35.7
YOLOv8n	63.5	44.9	39.2	24.8	6.0	33.2
YOLOv10n	61.0	44.2	37.5	23.3	5.5	30.9
YOLOv11n	65.6	56.6	56.7	32.6	9.3	26.6
IRE-YOLO	67.6	59.0	61.5	35.1	5.9	30.5

The results of the comparative experiments in Table 3 show that the proposed WI-YOLO model performs excellently in multiple key metrics, and its comprehensive performance outperforms that of current mainstream lightweight detection models. In terms of detection accuracy, WI-YOLO achieves 61.5% in mAP@50 and 35.1% in mAP@50:95, which are significantly 4.8 and 2.5 percentage points higher than those of the baseline model YOLOv11n, respectively. Meanwhile, the model's precision (P) and recall (R) are increased to 67.6% and 59.0%, respectively—indicating that while maintaining high classification accuracy, its ability to identify positive samples is also significantly enhanced.

In terms of model efficiency, WI-YOLO demonstrates an excellent balance of comprehensive performance. Its model size is only 5.9 MB, which represents a 36.6% reduction compared to YOLOv11n (9.3 MB), demonstrating good lightweight char-

acteristics. Of particular note, WI-YOLO achieves an inference speed of 30.5 FPS, which is not only higher than the 26.6 FPS of YOLOv11n but also outperforms peer models such as YOLOv10n and YOLOv8n. This achieves the dual improvement of accuracy and speed.

Compared with other lightweight variants of the YOLO series, WI-YOLO maintains a competitive model size while leading significantly in all detection accuracy metrics. Specifically, in terms of mAP@50, WI-YOLO has advantages of 21.9, 22.3, and 24.0 percentage points over YOLOv5n, YOLOv8n, and YOLOv10n, respectively. This fully demonstrates the effectiveness of its structural improvements. This excellent balance among accuracy, speed, and model complexity makes WI-YOLO particularly suitable for deployment on UAV embedded platforms with limited computing resources. In summary, WI-YOLO significantly reduces model complexity while comprehensively improving detection performance and inference efficiency, exhibiting strong competitiveness and practical application value among lightweight object detection models.

3.5. Visualization Results

To systematically evaluate the performance of the improved algorithm in real aerial environments, this study selected four types of typical complex scenarios from the DOTA dataset for visual experimental analysis, including dense small targets, variable lighting conditions, significant scale differences, and low-light environments. **Figure 4** and **Figure 5** present the comparison of detection results among the original images, the baseline YOLOv11n model, and the WI-YOLO algorithm, where red bounding boxes clearly mark the key difference regions.

The experimental results show that WI-YOLO exhibits excellent adaptability in all types of challenging scenarios. When facing densely distributed small targets, the algorithm significantly reduces missed detections and maintains a high recall rate even in regions with severe target overlap. Under conditions of drastic lighting changes, the model demonstrates superior feature preservation capability, effectively filtering out background interference and focusing on discriminative features. When there are targets with significant size differences in the same scene, WI-YOLO achieves accurate recognition of both tiny targets and large entities through the multi-scale feature fusion mechanism. In low-light environments, the algorithm effectively overcomes the impact of insufficient visibility on detection performance by virtue of its enhanced feature extraction capability.

Notably, in nighttime scenarios with unstable lighting conditions, WI-YOLO shows more stable detection performance compared to the baseline model. Comprehensive visual analysis indicates that the improved algorithm proposed in this study maintains a stable performance advantage across different types of complex aerial scenarios, particularly excelling in addressing lighting variations and scale diversity. This verifies its practical value and robustness in real-world environments.

Figure 4. Comparison of YOLOv11 and WI-TOLO Detection Results.

Figure 5. Comparison of YOLOv11 and IRE-TOLO Detection Result Diagrams.

4. Conclusion

To address the prevalent issues in UAV aerial images—including high difficulty in small target detection, strong background interference, significant target scale differences, and limited computing resources—this study improved the YOLOv11n model from two dimensions: feature enhancement and model lightweighting, and proposed an efficient aerial object detection algorithm, WI-YOLO. While maintaining high inference efficiency, this algorithm significantly improves detection accuracy and optimizes model complexity.

Specifically, this study first introduced the improved Efficient Multi-Scale Attention (iEMA) mechanism into the backbone network. By fusing standard deviation and local contrast metrics, the iEMA mechanism enhances the model's feature discrimination capability for low-contrast regions, thereby effectively improving target recognition accuracy in complex backgrounds. Second, a wavelet transform-based convolutional module (WTConv) was designed. Through a multi-level frequency domain decomposition and reconstruction mechanism, WTConv expands the recep-

tive field while preserving the low-frequency structural information and high-frequency detailed features of images, significantly improving the model's adaptability to multi-scale targets.

Experimental results on the DOTA dataset show that WI-YOLO outperforms mainstream lightweight detection models in key metrics such as mAP@50, mAP@50:95, and recall. Meanwhile, the model's parameter count is reduced to 5.9 MB, and the inference speed is increased to 30.5 FPS, achieving the coordinated optimization of detection accuracy and computational efficiency. However, the algorithm still exhibits a certain degree of missed detections in extreme scenarios (e.g., extremely dense targets or severe occlusions). Future research will focus on improving the model's generalization ability in complex scenarios and exploring its practical application potential in tasks such as dynamic inspection and real-time tracking.

References

[1] TIAN M ,CUI M ,CHEN Z , et al. MFR-YOLOv10: Object detection in UAV-taken images based on multilayer feature reconstruction network[J].Chinese Journal of Aeronautics,2025,38(11):103456-103456.DOI:10.1016/J.CJA.2025.103456.

[2] Zhang J ,Gao M ,Song L , et al. REA-YOLO for small object detection in UAV aerial images[J].The Journal of Supercomputing,2025,81(14):1332-1332.DOI:10.1007/S11227-025-07836-0.

[3] Yan Y ,Zhang L . A Review of YOLO Series Algorithms in Object Detection for UAV Aerial Images[J].International Core Journal of Engineering,2025,11(9):74-92.DOI:10.6919/ICJE.202509_11(9).0009.

[4] Zhou H ,Yin W ,Sun K , et al. FO-YOLO for small object detection in drone aerial imagery[J].The Journal of Supercomputing,2025,81(12):1208-1208.DOI:10.1007/S11227-025-07688-8.

[5] Bian Y ,Teng H ,Lv Q , et al. YOLO-AAHU: An adaptive multi-scale small target and occluded target detection algorithm for drone aerial photography scenarios[J].Engineering Research Express,2025,7(3):035226-035226.DOI:10.1088/2631-8695/ADEEFE.

[6] Ying Z ,Yisheng H .A Survey of SAR Image Target Detection Based on Convolutional Neural Networks[J].Remote Sensing,2022,14(24):6240-6240.

[7] Yihunie S A ,E. J B .A Survey on Deep-Learning-Based LiDAR 3D Object Detection for Autonomous Driving[J].Sensors,2022,22(24):9577-9577.

[8] Bao D ,Hao L .Survey of Target Detection Based on Neural Network[J].Journal of Physics: Conference Series,2021,1952(2):

[9] Kang T ,Yiquan W ,Fei Z .Recent advances in small object detection based on deep learning: A review[J].Image and Vision Computing,2020,97(prepublish):103910-103910.

[10] Zhao Z ,Zheng P ,Xu S , et al.Object Detection With Deep Learning: A Review.[J].IEEE Trans. Neural Netw. Learning Syst.,2019,30(11):3212-3232.